

REMOTE SENSING IMAGE REPRESENTATION BASED ON HIERARCHICAL HISTOGRAM PROPAGATION

*Jefersson A. dos Santos^{1,2}, Otávio A. B. Penatti¹, Ricardo da S. Torres¹,
Philippe-H. Gosselin², Sylvie Philipp-Foliguet², Alexandre X. Falcão¹*

¹RECOD Lab – Institute of Computing, University of Campinas, Brazil

²ETIS, CNRS, ENSEA, University of Cergy-Pontoise, France

ABSTRACT

Many methods have been recently proposed to deal with the large amount of data provided by high-resolution remote sensing technologies. Several of these methods rely on the use of image segmentation algorithms for delineating target objects. However, a common issue in geographic object-based applications is the definition of the appropriate data representation scale, a problem that can be addressed by exploiting multiscale segmentation. The use of multiple scales, however, raises new challenges related to the definition of effective and efficient mechanisms for extracting features. In this paper, we address the problem of extracting histogram-based features from a hierarchy of regions for multiscale classification. The strategy, called *H-Propagation*, exploits the existing relationships among regions in a hierarchy to iteratively propagate features along multiple scales. The proposed method speeds up the feature extraction process and yields good results when compared with global low-level extraction approaches.

Index Terms— Histogram propagation, image representation, multiscale analysis, GEOBIA.

1. INTRODUCTION

Remote sensing images are often used as data source for land cover studies in many applications [1]. A common problem in these applications is the definition of the data representation scale (the size of the segmented regions or block of pixels) [2, 3]. To address the segmentation scale problem, several approaches have been proposed considering multiscale analysis for applications that handle remote sensing images [1, 2, 4–6]. In these approaches, the feature extraction at various segmentation scales is an essential step. However, depending on the strategy, the extraction can be a very costly process. If we apply the same feature extraction algorithm for all regions of different segmentation scales,

for example, the pixels in the image would need to be accessed at least once for each scale.

Aiming at addressing this issue, we have recently proposed an strategy called *BoW-Propagation* [7], which exploits the bag-of-words concept [8, 9] to iteratively propagate texture features along the hierarchy from the finest regions to the coarsest ones. That strategy speeds up the extraction process by avoiding the computation of low-level features from all regions of the hierarchy.

In this paper, we propose an approach, called *H-Propagation* to extract histogram-based features from a hierarchy of segmented regions. This approach, which is an extension of the *BoW-Propagation*, relies on processing only the image pixels in the base of the hierarchy (the finest region scale). The features are quickly propagated to the upper scales by exploiting the hierarchical association among regions at different scales. The features are then propagated to the other scales. At the end, all regions in the hierarchy are represented by a histogram.

2. BOW-PROPAGATION

The *BoW-Propagation* [7], which is based on the bag-of-visual-word model [10], is designed to extract features from a hierarchy of segmented regions [11]. It relies on processing only the image pixels in the base of the hierarchy (the finest regions scale). It exploits the hierarchical relationship among regions at different scales to compute the features.

The strategy starts by creating a visual dictionary based on low-level features extracted from the pixel level (the base of the hierarchy), as shown in Figure 1. The low-level feature space is quantized, creating the visual words, and each region in the base of the hierarchy is described according to that dictionary. The features are then propagated to the other scales. At the end, all regions in the hierarchy will be represented by a bag of visual words.

Figure 2 illustrates a schema to represent a segmented region by using dense sampling through a bag of words. The low-level features extracted from the

This work was supported by CAPES (592/08 and BEX 5224101), FAPESP (Grants 2008/58528-2, 2010/52113-5, 2012/16253-2, and 2012/18768-0), and CNPq (303673/2010-9, 484254/2012-0, and 306580/2012-8).

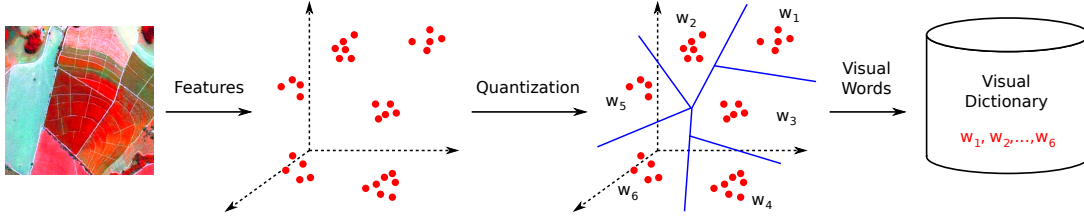


Fig. 1. Construction of a visual dictionary to describe a remote sensing image. The features are extracted from groups of pixels (e.g., tiles or segmented regions), the feature space is quantized so that each cluster corresponds to a visual word w_i .

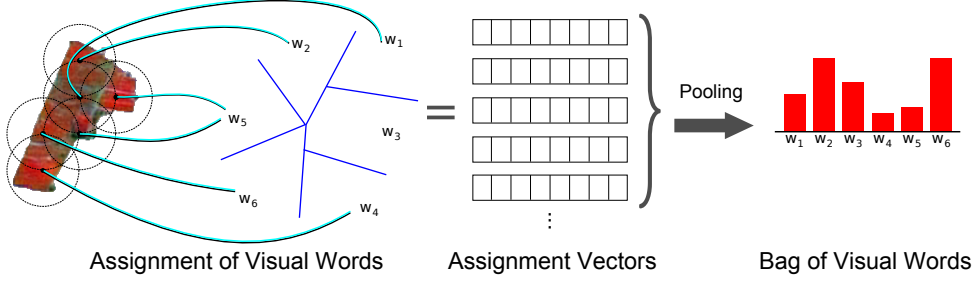


Fig. 2. Representation of a segmented region based on a visual dictionary with dense sampling feature extraction.

internal points are assigned to visual words and combined by a pooling function.

3. THE PROPOSED HISTOGRAM FEATURE PROPAGATION

The histogram propagation (*H-Propagation*) consists in estimating the feature histogram representation of a region R , given the low-level histograms extracted from the R subregions $\Gamma(R)$. It is an extension of the *BoW-propagation* in the sense it propagates any kind of low-level features based on histograms from fine scales to the coarsest ones.

Let P_{λ_x} and P_{λ_y} be partitions obtained from the hierarchy H at the scales $\lambda_x \in \lambda_y$, respectively. We consider that $P_{\lambda_x} > P_{\lambda_y}$, i.e., P_{λ_x} is coarser than P_{λ_y} . Let $R \in P_{\lambda_x}$ be a region from the partition P_{λ_x} . We call *subregion* of R the region $\hat{R} \in P_{\lambda_y}$ such that $\hat{R} \subseteq R$.

The set $\Gamma(R)$, which is composed of the subregions of R in the partition P_{λ_y} , is given by $\Gamma(R) = \{\forall \hat{R} \in P_{\lambda_y} | p \in R \cap p \in \hat{R}\}$, where p is a pixel. The set of subregions of R in a finer scale are all the regions \hat{R} that have all pixels inside \hat{R} and inside R .

The principle of *H-propagation* is to compute the feature histogram h_R , which describes region R , by combining the histograms of subregions $\Gamma(R)$:

$$h_R = f\{h_{\hat{R}_c} \mid \hat{R}_c \in \Gamma(R)\} \quad (1)$$

where f is a combination function.

Algorithm 1 presents the proposed H-propagation. The first step is to extract low-level features from the regions in the finest scale λ_1 (line 1). The “propagation loop” is responsible for propagating the features to other scales (lines 2 to 6). For all regions R from a partition P_{λ_x} , the histogram h_R is computed based on the $\Gamma(R)$ histograms, which is defined by Equation 1 (line 4).

Algorithm 1 H-Propagation

- 1 Extract low-level feature histograms from the regions in the finest scale λ_1
 - 2 For $i \leftarrow 2$ to n do
 - 3 For all $R \in P_{\lambda_i}$ do
 - 4 Compute the histogram h_R based on the $\Gamma(R)$ histograms
 - 5 End for
 - 6 End for
-

Figure 3 illustrates an example of the use of the combination function f to compute the histogram h_r of a region r . The region $r \in P_{\lambda_2}$ is composed of the set of subregions $\Gamma(r) = \{a, b, c\}$ at the scale λ_1 . Figure 3 (a) illustrates, in gray, the region r and its subregions $\Gamma(r)$ in the hierarchy of regions. In Figure 3 (b), the histogram h_r is computed based on the function f : $h_r = f(h_a, h_b, h_c)$. Figure 4 illustrates the computation of h_r by using the *max* operator as combination function.

H-propagation does not quantize the low-level fea-

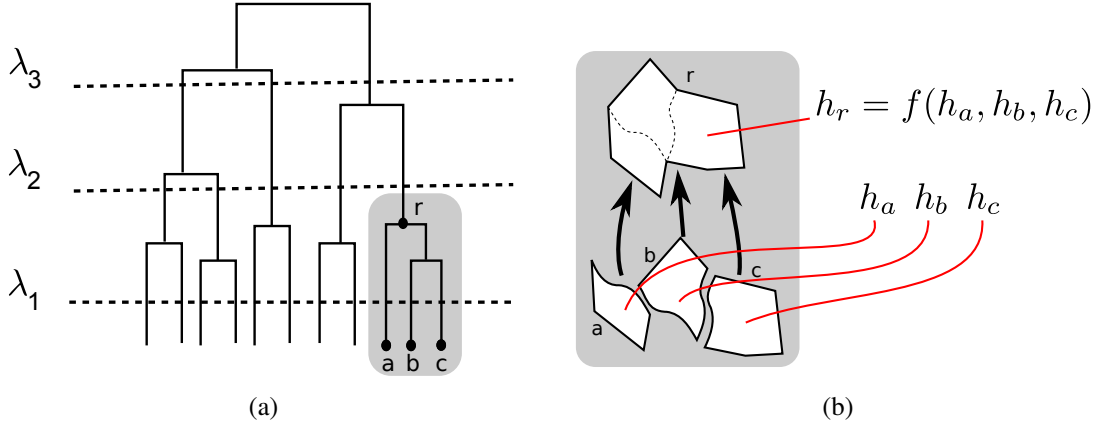


Fig. 3. Computing the histogram h_r of region r by combining the histograms h_a , h_b , and h_c from the subregions a , b , and c .

ture space to create a visual dictionary. Another difference, when compared with the BoW-propagation, is that H-propagation propagates histogram bins instead of the probabilities of visual words. BoW-propagation is suitable for propagating low-level local features. H-propagation, in turn, is designed only for global descriptors based on histogram representations.

An important issue is the definition of the propagation function f in the case of low-level histograms. Contrarily to the propagation of visual words, we use the average function instead of the max function. It is expected that with the average propagation, the quality of the histograms will be the same as that performed by the extraction directly from the pixels at all scales of the hierarchy.

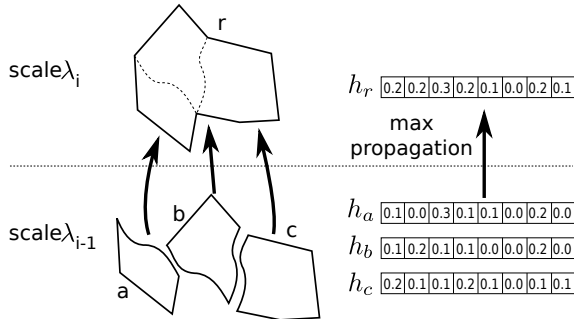


Fig. 4. Feature propagation example using a *max* pooling operation.

4. SIMULATIONS

We have used two different datasets in our experiments. We refer to the dataset according to the target regions: COFFEE and URBAN. We carried out experiments with ten different combinations of the nine

subimages for each dataset (three for training, three for validation, and three for testing). The COFFEE dataset is a SPOT image with $2.3m$ of spatial resolution from Monte Santo de Minas, Brazil. The URBAN dataset is a Quickbird image with $0.6m$ of spatial resolution from Campinas, Brazil. More details about the datasets can be found in [1]. We define the scales according to the principle of dichotomic cuts proposed in [11]. The higher the index, the coarser is the scale. The scale λ_0 is the pixel level. We used linear SVMs to obtain the classification results. It is evaluated by computing the overall accuracy, kappa index, and tau index for the classified images [12].

We have tested the proposed approaches for color feature propagation. We have selected BIC descriptor [13] since it produced good results in previous works considering multiscale classification [4, 14, 15]. We compare the propagation approaches against BIC low-level feature extraction [13].

BIC BoW-Propagation was computed by using: *max* pooling function, dictionary size of 10^3 words, and soft assignment ($\sigma = 0.1$). We have extracted low-level features from a dense sampling by overlapping squares with 4×4 pixels. BIC H-Propagation, in turn, was computed by using the *avg* pooling function.

Table 1 presents the classification using the BIC descriptor with BoW-Propagation, Histogram Propagation, and low-level extraction (Global Descriptor) for the COFFEE dataset.

Concerning the results for the COFFEE dataset, H-Propagation and the Global Descriptor present the same overall accuracy (around 80%). The same can be observed for *kappa* and *tau* indexes. BoW-Propagation achieved results slightly worse than the other two approaches for the three computed measures. Regarding the URBAN dataset, H-Propagation and Global Descriptor obtained the same overall accuracy, Kappa, and Tau ($\sim 70\%$, 0.31, and 0.47,

Table 1. Classification results for the BIC descriptor using BoW-Propagation, H-Propagation, and global feature extraction for the COFFEE and the URBAN datasets at segmentation scale λ_3 .

Method	COFFEE			URBAN		
	O.A. (%)	Kappa (κ)	Tau (τ)	O.A. (%)	Kappa (κ)	Tau (τ)
BoW-Propagation	73.41 \pm 2.76	0.25 \pm 0.03	0.36 \pm 0.02	67.03 \pm 2.65	0.26 \pm 0.03	0.47 \pm 0.02
H-Propagation	79.97 \pm 1.76	0.46 \pm 0.02	0.54 \pm 0.02	69.86 \pm 4.76	0.31 \pm 0.05	0.47 \pm 0.04
Global Descriptor	80.07 \pm 1.81	0.47 \pm 0.02	0.54 \pm 0.02	69.63 \pm 3.33	0.31 \pm 0.04	0.47 \pm 0.03

respectively). The BoW-Propagation approach yields slightly worse results than the other methods concerning overall accuracy and Kappa index. The Tau index was the same (0.47).

5. CONCLUSION

The proposed H-propagation approach revealed be suitable for saving time on feature extraction from a hierarchy of segmented regions. Regarding color features, BOW-Propagation seems to be promising, but it requires many setup parameters. However, H-Propagation shows that it is possible to compute low-level features based only on the extracted hierarchy. Furthermore, these features can be propagated without losses in terms of representation quality.

Future work considers the study the use of the H-propagation method in other applications, such as content-based image retrieval and multimedia forensics.

6. REFERENCES

- [1] J. A. dos Santos, P.H. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. X. Falcão, "Interactive multiscale classification of high-resolution remote sensing images," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2013, To appear.
- [2] Y. Tarabalka, J.C. Tilton, J.A. Benediktsson, and J. Chanussot, "A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 1, pp. 262–272, feb. 2012.
- [3] J. A. dos Santos, C. D. Ferreira, R. da S. Torres, M. A. Gonçalves, and R. A. C. Lamparelli, "A relevance feedback method based on genetic programming for classification of remote sensing images," *Information Sciences*, vol. 181, no. 13, pp. 2671–2684, 2011.
- [4] J. A. dos Santos, P.H. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. X. Falcão, "Multiscale classification of remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 10, pp. 3764–3775, 2012.
- [5] J. Munoz-Mari, D. Tuia, and G. Camps-Valls, "Semisupervised classification of remote sensing images with active queries," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 10, pp. 3751–3763, oct. 2012.
- [6] A. Alonso-González, S. Valero, J. Chanussot, C. López-Martínez, and P. Salembier, "Processing multidimensional sar and hyperspectral images with binary partition tree," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 723–747, 2013.
- [7] J. A. dos Santos, O. A. B. Penatti, R. da S. Torres, P-H. Gosselin, S. Philipp-Foliguet, and A. X. Falcão, "Improving texture description in remote sensing image multi-scale classification tasks by using visual words," in *International Conference on Pattern Recognition*, Tsukuba, Japan, November 2012, pp. 3090–3093.
- [8] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," *Conference on Computer Vision and Pattern Recognition*, pp. 2559–2566, 2010.
- [9] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J-M Geusebroek, "Visual word ambiguity," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1271–1283, 2010.
- [10] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," in *International Conference on Computer Vision*, 2003, pp. 1470–1477 vol.2.
- [11] L. Guigues, J. Cocquerez, and H. Le Men, "Scale-sets image analysis," *International Journal of Computer Vision*, vol. 68, pp. 289–317, 2006.
- [12] Zhenkui Ma and Roland Redmond, "Tau coefficients for accuracy assessment of classification of remote sensing data," *Photogrammetric Engineering and Remote Sensing*, vol. 61, no. 4, pp. 439–453, 1995.
- [13] R. de O. Stehling, M. A. Nascimento, and A. X. Falcão, "A compact and efficient image retrieval approach based on border/interior pixel classification," in *CIKM*, New York, NY, USA, 2002, pp. 102–109.
- [14] J. A. dos Santos, F. A. Faria, R. da S. Torres, A. Rocha, P-H. Gosselin, S. Philipp-Foliguet, and A. X. Falcão, "Descriptor correlation analysis for remote sensing image multi-scale classification," in *International Conference on Pattern Recognition*, Tsukuba, Japan, November 2012, pp. 3078–3081.
- [15] J. A. dos Santos, O. A. B. Penatti, and R. da S. Torres, "Evaluating the potential of texture and color descriptors for remote sensing image retrieval and classification," in *The International Conference on Computer Vision Theory and Applications*, Angers, France, May 2010, pp. 203–208.